

# Penerapan Algoritma C4.5 dalam Menentukan Faktor yang Dapat Meningkatkan Kemampuan Bahasa Inggris pada Mahasiswa

Septri Wanti Siahaan<sup>1</sup>, Kristin D R Sianipar<sup>2</sup>, P.P.P.A.N.W Fikrul Ilmi R.H Zer<sup>3</sup>, Dedy Hartama<sup>4</sup>

Program Studi Teknik Informatika  
STIKOM Tunas Bangsa Pematangsiantar  
Pematangsiantar, Indonesia

e-mail: <sup>1</sup>septriwanti26@gmail.com, <sup>2</sup>kristinsianipar7@gmail.com, <sup>3</sup>fikrulilmizer@gmail.com, <sup>4</sup>dedyhartama@amiktunasbangsa.ac.id

Diajukan: 14 Juli 2020; Direvisi: 25 September 2020; Diterima: 30 September 2020

## Abstrak

Negara-negara di dunia memiliki bahasa tersendiri. Bahasa yang paling dominan digunakan adalah Bahasa Inggris. Oleh karena itu, Bahasa Inggris merupakan Bahasa Internasional. Bahasa internasional adalah bahasa yang digunakan agar dapat berkomunikasi dengan orang-orang di dunia. Indonesia merupakan salah satu negara yang menerapkan Bahasa Inggris sebagai mata pelajaran atau mata kuliah wajib yang dipelajari oleh siswa/i dan mahasiswa/i. Mahasiswa/i wajib mengetahui bahwa mempelajari Bahasa Inggris sangat dibutuhkan untuk meningkatkan kompetensi di masa depan. Di era globalisasi saat ini, Bahasa Inggris dapat memberikan dampak yang mendominasi seperti komunikasi, pekerjaan, studi, dan lainnya. Namun, seperti yang semua orang tahu bahwa, mempelajari Bahasa Inggris tidaklah mudah. Banyak mahasiswa/i yang mendaftar kelas Bahasa Inggris untuk meningkatkan kemampuan Bahasa Inggris. Namun, ada juga mahasiswa/i yang tidak mendaftar kelas Bahasa Inggris tetapi, memiliki kemampuan Bahasa Inggris yang bagus. Dalam menyelesaikan penelitian ini, penulis menerapkan Algoritma yang ada pada Data Mining yaitu Algoritma C4.5. Hasil dari penelitian ini adalah faktor "Mendengar dari Lingkungan" mendapatkan gain tertinggi pada iterasi ke-1. Kesimpulan dari penelitian ini adalah bahwa faktor yang mempengaruhi dalam meningkatkan kemampuan Bahasa Inggris pada mahasiswa adalah "Mendengar dari Lingkungan". Penelitian ini memiliki tujuan untuk menentukan faktor yang meningkatkan kemampuan Bahasa Inggris pada mahasiswa.

**Kata kunci:** Algoritma C4.5, Bahasa Inggris, Data mining, Daya tangkap, Peningkatan.

## Abstract

The countries of the world have their own languages. The most dominant language used is English. Therefore, English is an international language. International language is the language used in order to communicate with people in the world. Indonesia is one of the countries that applies English as a compulsory subject or subject to be studied by students and students. Students must know that learning English is needed to improve competence in the future. In the current era of globalization, English can have a dominating impact such as communication, work, study, and others. However, as everyone knows, learning English is not easy. Many students enroll in English classes to improve their English skills. However, there are also students who do not register for English classes but have good English skills. In completing this research, the authors apply the existing algorithm in data mining, namely the C4.5 algorithm. The result of this research is the "Hearing from the Environment" factor gets the highest gain in the 1st iteration. The conclusion of this study is that the factor that affects students' English skills improvement is "Hearing from the Environment". This study aims to determine the factors that improve students' English skills.

**Keywords:** C4.5 algorithm, English, Data mining, Comprehension, Enhancement.

## 1. Pendahuluan

Negara-negara di dunia memiliki bahasa yang digunakan dalam berkomunikasi dengan orang lain. Di Indonesia menggunakan bahasa persatuan yaitu Bahasa Indonesia. Namun, jika ingin berkomunikasi

dengan orang asing dari negara lain, pasti akan kesulitan untuk mengerti apa yang mereka katakan. Hal itu dikarenakan manusia menggunakan bahasa yang mereka miliki dan begitu pun dengan orang asing, sehingga tidak ada titik temu dalam komunikasi. Oleh karena itu, Bahasa Inggris digunakan sebagai bahasa global dalam berkomunikasi. Sehingga, orang-orang di dunia dapat berkomunikasi dan mengerti apa yang diucapkan. Salah satu teknologi yang dapat dipakai untuk meningkatkan dalam pembelajaran adalah *computer* [1].

Pada Negara Indonesia, semua orang mempelajari Bahasa Inggris sudah sejak dari Sekolah Dasar (SD) ataupun ada juga yang telah mempelajarinya dari taman kanak-kanak. Di Pendidikan Tinggi sekalipun menerapkan untuk memberikan mata kuliah Bahasa Inggris kepada mahasiswa/i meskipun *output* dari studi tidak berhubungan dengan Bahasa Inggris. Dari hal tersebut, dapat disimpulkan bahwa betapa pentingnya kemampuan dalam berbahasa asing, terkhususnya Bahasa Inggris sebagai salah satu faktor kesuksesan pada akademik seseorang ataupun untuk mendapatkan karier bagus dalam pekerjaan [2]. Namun, walaupun telah mempelajari Bahasa Inggris sudah sejak dari lama, masih banyak mahasiswa yang merasa kesulitan dalam mempelajari Bahasa Inggris.

Mahasiswa kesulitan dalam mempelajari Bahasa Inggris. Kesulitan yang dapat dialami seperti, kesulitan dalam membaca, menulis, mendengarkan, dan lain-lain. Terkadang ada dari mahasiswa yang masih merasa tidak percaya diri dalam mengaplikasikan bahasa tersebut dalam kehidupan sehari-hari. Alasannya adalah karena takut diolok dan dianggap pamer bahwa mampu dalam Bahasa Inggris.

Adapun juga masalah yang sering dialami mahasiswa yaitu kesulitan dalam membaca kosakata dan menerapkan *grammar* dalam Bahasa Inggris. Peran pengajar sangat diperlukan untuk dapat meningkatkan kemampuan Bahasa Inggris pada mahasiswa.

Dengan adanya permasalahan tersebut, penulis berinisiatif untuk menentukan faktor apa yang dapat meningkatkan kemampuan mahasiswa dalam Bahasa Inggris. Ada banyak cabang dari ilmu komputer yang dapat digunakan untuk menyelesaikan permasalahan yang sifatnya kompleks. Cabang ilmu komputer tersebut adalah *Artificial Intelligence* seperti *datamining* [3]–[9]. Berdasarkan permasalahan di atas, penulis memanfaatkan Data Mining yaitu klasifikasi dalam menyelesaikan permasalahan untuk meningkatkan kemampuan Bahasa Inggris pada mahasiswa. Algoritma pada Data Mining yang digunakan adalah Algoritma C4.5 [10]. Tujuan dilakukannya penelitian ini adalah untuk menentukan faktor yang dapat meningkatkan kemampuan Bahasa Inggris pada mahasiswa.

## 2. Metode Penelitian

### 2.1. Membuat Dataset

*Dataset* diperoleh dengan memberikan kuesioner yang dibuat dari Google Form. Kuesioner diberikan kepada mahasiswa untuk melengkapi penyelesaian penelitian ini.

Tabel 1. Data responden.

No	K1	K2	K3	K4	Kemampuan Bahasa Inggris
1.	3	4	5	1	Mampu
2.	4	3	1	2	Mampu
3.	1	5	5	3	Mampu
4.	3	5	2	2	Tidak Mampu
5.	5	4	4	3	Mampu
6.	4	1	2	2	Tidak Mampu
7.	5	4	5	5	Mampu
8.	4	5	5	3	Mampu
9.	3	4	3	1	Tidak Mampu
10.	1	3	2	3	Mampu
11.	4	2	3	3	Mampu
12.	5	4	1	5	Tidak Mampu
13.	2	2	3	3	Tidak Mampu
14.	4	4	5	5	Mampu
15.	5	5	3	5	Mampu
16.	1	4	3	2	Tidak Mampu
17.	4	1	4	4	Tidak Mampu
18.	5	2	2	3	Tidak Mampu
19.	2	5	5	1	Tidak Mampu
20.	4	4	5	4	Mampu
21.	4	3	1	2	Mampu
22.	1	5	5	3	Mampu
23.	3	5	2	2	Tidak Mampu
24.	4	4	5	5	Mampu

No	K1	K2	K3	K4	Kemampuan Bahasa Inggris
25.	5	5	3	5	Mampu
26.	1	4	3	2	Tidak Mampu
27.	4	1	4	4	Tidak Mampu
28.	5	2	2	3	Tidak Mampu
29.	2	5	5	1	Tidak Mampu
30.	1	5	5	3	Mampu
31.	3	5	2	2	Tidak Mampu
32.	5	4	4	3	Mampu
33.	3	4	3	1	Tidak Mampu
34.	1	3	2	3	Mampu
35.	4	2	3	3	Mampu
36.	5	4	1	5	Tidak Mampu
37.	1	4	3	2	Tidak Mampu
38.	4	1	4	4	Tidak Mampu
39.	4	1	2	2	Tidak Mampu
40.	5	4	5	5	Mampu
41.	4	5	5	3	Mampu
42.	3	4	3	1	Tidak Mampu
43.	1	3	2	3	Mampu
44.	3	4	5	1	Mampu
45.	4	3	1	2	Mampu
46.	2	5	5	1	Tidak Mampu
47.	5	4	1	5	Tidak Mampu
48.	2	2	3	3	Tidak Mampu
49.	4	4	5	5	Mampu
50.	5	5	3	5	Mampu
51.	2	5	5	1	Tidak Mampu
52.	4	4	5	4	Mampu
53.	4	3	1	2	Mampu
54.	5	4	1	5	Tidak Mampu
55.	2	2	3	3	Tidak Mampu
56.	5	4	4	3	Mampu
57.	4	1	2	2	Tidak Mampu
58.	5	4	5	5	Mampu
59.	4	1	2	2	Tidak Mampu
60.	5	4	5	5	Mampu
61.	4	3	1	2	Mampu
62.	1	5	5	3	Mampu
63.	3	5	2	2	Tidak Mampu
64.	5	4	4	3	Mampu
65.	1	4	3	2	Tidak Mampu
66.	4	1	4	4	Tidak Mampu
67.	4	1	2	2	Tidak Mampu
68.	2	5	5	1	Tidak Mampu
69.	4	4	5	4	Mampu
70.	4	3	1	2	Mampu

Tabel 1 merupakan data dari 70 responden. Pada Tabel 1 dibuat 4 faktor yang mempengaruhi kemampuan Bahasa Inggris pada mahasiswa, yaitu: Membaca Referensi (K1), Mendengar dari Lingkungan (K2), Mempraktikkan dalam Lingkungan (K3), dan Memanfaatkan Teknologi (K4). *Output* yang diinginkan yaitu Mampu dan Tidak Mampu. Dengan sub kriteria sebagai berikut:

- 5 = Sangat Setuju
- 4 = Setuju
- 3 = Cukup Setuju
- 2 = Kurang Setuju
- 1 = Sangat Kurang Setuju

## 2.2. Menerapkan Algoritma yang Digunakan

Penelitian ini menggunakan Algoritma C4.5 dalam menentukan faktor yang mempengaruhi mahasiswa dalam meningkatkan kemampuan Bahasa Inggris mahasiswa. Algoritma C4.5 adalah algoritma yang terdapat pada teknik pengklasifikasian untuk menyelesaikan kasus atau pun masalah. Pohon keputusan (*decision tree*) merupakan dasar dari Algoritma C4.5. Algoritma C4.5 adalah salah satu algoritma induksi pohon keputusan yaitu ID3 (*Iterative Dichotomiser 3*) [11].

Adapun rumus yang dipakai dalam Algoritma C4.5 yaitu:

- Menghitung *entropy*

$$Entropi (S) = \sum_{j=1}^k - p_j \log_2 p_j \quad (1)$$

Variabel  $S$  merupakan himpunan (*dataset*) kasus dan  $k$  merupakan banyaknya partisi dari *dataset*.

- Menghitung *gain*

$$Gain (A) = Entropi (S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi (S_i) \quad (2)$$

$S$  merupakan himpunan (*dataset*) kasus,  $A$  merupakan atribut,  $N$  merupakan banyaknya partisi atribut,  $|S_i|$  menunjukkan banyaknya jumlah kasus partisi ke- $i$ ,  $|S|$  merupakan banyaknya jumlah kasus  $S$ .

Kelebihan dari menggunakan Algoritma C4.5 adalah dapat membuat pohon keputusan (*decision tree*) sehingga menjadi efisien, di mana *decision tree* menangani atribut tipe diskrit dan tipe diskrit-numerik, mudah untuk diinterpretasikan dan memiliki tingkat akurasi yang dapat diterima [12]. Kelemahan Algoritma C4.5 salah satunya adalah terdapat pada skalabilitas yaitu *data training* hanya dapat digunakan dan disimpan secara keseluruhan pada waktu yang bersamaan di memori.

### 2.3. Melakukan Hitungan Manual

Pada penelitian ini dilakukan perhitungan manual dengan menggunakan rumus yang telah ada dan dilakukan dengan menggunakan Microsoft Office Excel. Pada saat melakukan hitungan manual harus dilakukan dengan teliti agar didapatkan hasil yang benar dan sesuai dengan *software* Rapidminer.

- Iterasi ke-1

$$\begin{aligned} Entrophy (Total) &= \left( -\left(\frac{34}{70}\right) \times \log_2 \left(\frac{34}{70}\right) + -\left(\frac{36}{70}\right) \times \log_2 \left(\frac{36}{70}\right) \right) \\ &= (0,4857 \times 1,0365) + (0,5143 \times 0,9593) \\ &= 0,9994 \end{aligned}$$

$$\begin{aligned} Gain (S, Membaca Referensi) &= 0,9994 - \left( \left(\frac{17}{70} \times 0,9367\right) + \left(\frac{25}{70} \times 0,9427\right) + \right. \\ &\quad \left. \left(\frac{9}{70} \times 0,7642\right) + \left(\frac{8}{70} \times 0\right) + \left(\frac{11}{70} \times 0,9457\right) \right) \\ &= 0,1884 \end{aligned}$$

- Iterasi ke-2

$$\begin{aligned} Entrophy (Total) &= \left( -\left(\frac{9}{18}\right) \times \log_2 \left(\frac{9}{18}\right) + -\left(\frac{9}{18}\right) \times \log_2 \left(\frac{9}{18}\right) \right) \\ &= 1 \end{aligned}$$

$$\begin{aligned} Gain (S, Membaca Referensi) &= 1 - \left( \left(\frac{3}{18} \times 0\right) + \left(\frac{2}{18} \times 0\right) + \left(\frac{4}{18} \times 0\right) + \left(\frac{5}{18} \times 0\right) + \right. \\ &\quad \left. \left(\frac{4}{18} \times 0\right) \right) \\ &= 1 \end{aligned}$$

### 2.4. Validasi

Proses validasi dilakukan dengan menggunakan *software* Rapidminer. Validasi dilakukan dengan mengambil sampel dari *dataset* yang dimiliki sebanyak 70 responden dari data kuesioner yang diberikan kepada mahasiswa dan kemudian data diuji dengan menggunakan *decision tree* (C4.5).

*Decision tree* memakai kinerja dalam bentuk pohon (*tree*) yang dimana pada masing-masing *node* yang dimiliki menjelaskan atribut yang ada, cabangnya menafsirkan nilai pada atribut yang ada dan *leaf* mempresentasikan kelas [13]. Pada pohon keputusan ada yang dikenal sebagai *root*. *Root* adalah *node* yang berada paling atas dari *decision tree*. *Decision tree* adalah teknik pengklasifikasian data yang paling dikenal untuk digunakan pada *data mining*. Pohon keputusan menggunakan pembangunan yang relatif cepat. *Output* dari jenis model yang dibuat untuk dimengerti secara mudah.

### 2.5. Data Mining

*Data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database* [14]. *Data mining* juga dikenal dengan istilah *pattern recognition* merupakan suatu

algoritma yang digunakan untuk pengolahan data guna menemukan pola yang tersembunyi dari data yang diolah. Data yang diolah kemudian menghasilkan suatu pengetahuan baru yang bersumber dari data.

*Data mining* memiliki dasar acuan seperti proses pencarian informasi yang sebelumnya tidak diketahui, di mana terdiri sekumpulan data-data besar. Tujuan dari *data mining* adalah menemukan, memperoleh pola atau hubungan yang akan memberikan manfaat. *Data mining* biasanya memiliki proses sebagai berikut:

- Deskripsi
- Prediksi
- Estimasi
- Klasifikasi
- *Clustering*

## 2.6. Decision Tree (Pohon Keputusan)

*Decision tree* memakai kinerja dalam bentuk pohon (*tree*) yang di mana pada masing-masing *node* yang dimiliki menjelaskan atribut yang ada, cabangnya menafsirkan nilai pada atribut yang ada dan *leaf* mempresentasikan kelas [13]. Pada pohon keputusan ada yang dikenal sebagai *root*. *Root* adalah *node* yang berada paling atas dari pohon keputusan. *Decision tree* adalah metode pengklasifikasian data yang paling dikenal untuk digunakan pada *data mining*. Pohon keputusan menggunakan pembangunan yang relatif cepat. *Output* dari jenis model yang dibuat mudah untuk dimengerti. Pohon keputusan terbagi 3 jenis *node*, yaitu sebagai berikut:

- a. *Root Node*, merupakan *node* yang berada paling atas. Pada *node* ini tidak memiliki *input* dan dapat juga tidak memiliki *output* atau memiliki *output* yang lebih dari satu.
- b. *Internal Node*, merupakan *node* yang bercabang. Pada *node* ini hanya memiliki 1 buah *input* dan memiliki *output* yaitu minimal 2.
- c. *Leaf Node* atau *Terminal Node*, merupakan *node* yang di akhir. Pada *node* ini hanya memiliki 1 *input* dan tidak memiliki *output*.

## 2.7. Algoritma C4.5

Secara garis besar, langkah-langkah yang dilakukan oleh Algoritma C.45 dalam membentuk pohon keputusan adalah sebagai berikut [15] :

*Input* : sampel *training*, label *training*, atribut

1. Membuat simpul akar untuk pohon yang dibuat
2. Jika semua sampel positif, berhenti dengan suatu pohon dengan satu simpul akar, beri tanda (+)
3. Jika semua sampel negatif, berhenti dengan suatu pohon dengan satu simpul akar, beri tanda (-)
4. Jika atribut kosong, berhenti dengan suatu pohon dengan satu simpul akar, dengan label sesuai nilai yang terbanyak yang ada pada label *training*
5. Untuk yang lain, Mulai
  - a. A ---- atribut yang mengklasifikasikan sampel dengan hasil terbaik (berdasarkan *gain ratio*)
  - b. Atribut keputusan untuk simpul akar ---- A
  - c. Untuk setiap nilai,  $v_i$ , yang mungkin untuk A
    - 1) Tambahkan cabang di bawah akar yang berhubungan dengan  $A = v_i$
    - 2) Tentukan sampel  $S_{v_i}$  sebagai subset dari sampel yang mempunyai nilai  $v_i$  untuk atribut A
    - 3) Jika sampel  $S_{v_i}$  kosong
      - i. Di bawah cabang tambahkan simpul daun dengan label = nilai yang terbanyak yang ada pada label *training*
      - ii. Yang lain tambah cabang baru di bawah cabang yang sekarang C4.5 (sampel *training*, label *training*, atribut – [A]).
  - d. Berhenti mengubah *tree* yang dihasilkan dalam beberapa *rule*. Jumlah *rule* sama dengan jumlah *path* yang mungkin dapat dibangun dari *root* sampai *leaf node*.

## 3. Hasil dan Pembahasan

Penelitian ini dilakukan dengan mengumpulkan data dari 70 responden dengan membuat 4 kriteria, yaitu: Membaca Referensi (K1), Mendengar dari Lingkungan (K2), Mempraktikkan dalam Lingkungan (K3), dan Memanfaatkan Teknologi (K4).

Dilakukan pengklasifikasian pada data responden. Perhitungan dilakukan dengan menggunakan *software* Microsoft Excel. Maka, selanjutnya melakukan perhitungan data menggunakan rumus yang ada pada Algoritma C4.5. Hasil dari perhitungan pada iterasi ke-1 dapat dilihat sebagai berikut:

Tabel 2. Hasil perhitungan iterasi ke-1.

		Jlh Kasus	Mampu	Tidak Mampu	Entropy	Gain
		(S)	(S1)	(S2)		
Node 1		70	36	34	0,999411	
Membaca Referensi	Sangat Setuju	17	11	6	0,936667	
	Setuju	25	16	9	0,942683	
	Cukup Setuju	9	2	7	0,764205	
	Kurang Setuju	8	0	8	0	
	Sangat Kurang Setuju	11	7	4	0,94566	
						0,188403
Mendengar dari Lingkungan	Sangat Setuju	18	9	9	1	
	Setuju	27	16	11	0,975119	
	Cukup Setuju	9	9	0	0	
	Kurang Setuju	7	2	5	0,863121	
	Sangat Kurang Setuju	9	0	9	0	
						0,279838
Mempraktikkan dari Lingkungan	Sangat Setuju	23	18	5	0,755375	
	Setuju	8	4	4	1	
	Cukup Setuju	15	5	10	0,918296	
	Kurang Setuju	14	3	11	0,749595	
	Sangat Kurang Setuju	10	6	4	0,970951	
						0,151526
Memanfaatkan Teknologi	Sangat Setuju	14	10	4	0,863121	
	Setuju	7	3	4	0,985228	
	Cukup Setuju	20	15	5	0,811278	
	Kurang Setuju	19	6	13	0,899744	
	Sangat Kurang Setuju	10	2	8	0,721928	
						0,149121

Pada Tabel 2 merupakan hasil perhitungan pada iterasi ke-1. Dapat dilihat bahwa faktor “Membaca Referensi” memiliki *gain* dengan nilai 0,188403, “Mendengar dari Lingkungan” memiliki *gain* dengan nilai 0,279838, “Mempraktikkan dari Lingkungan” memiliki *gain* dengan nilai 0,151526, “Memanfaatkan Teknologi” memiliki *gain* dengan nilai 0,149121. Oleh karena itu, faktor “Mendengar dari Lingkungan” mendapatkan *gain* tertinggi dengan nilai adalah 0,27984 dan nilai *entropy* yang tertinggi adalah “Sangat Setuju” dengan nilai adalah 1.

Dari Tabel 2, dikarenakan terdapat *root gain* tertinggi, maka dilanjutkan ke iterasi ke-2. Hasil perhitungan pada iterasi ke-2 dapat dilihat di bawah ini:

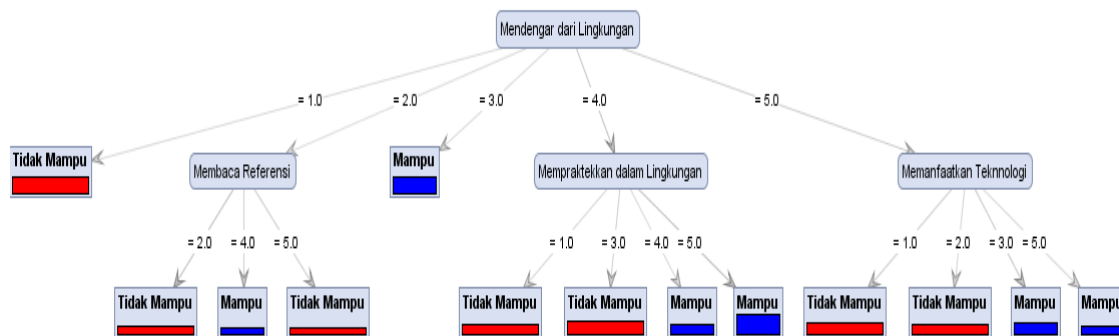
Tabel 3. Hasil perhitungan iterasi ke-2.

		Jlh Kasus	Mampu	Tidak Mampu	Entropy	Gain
		(S)	(S1)	(S2)		
Node 2		18	9	9	1	
Membaca Referensi	Sangat Setuju	3	3	0	0	
	Setuju	2	2	0	0	
	Cukup Setuju	4	0	4	0	
	Kurang Setuju	5	0	5	0	
	Sangat Kurang Setuju	4	4	0	0	
						1
Mempraktikkan dari Lingkungan	Sangat Setuju	11	6	5	0,9940302	
	Setuju	0	0	0	0	
	Cukup Setuju	3	3	0	0	
	Kurang Setuju	4	0	4	0	

		Jlh Kasus	Mampu	Tidak Mampu	Entropy	Gain
Memanfaatkan Teknologi	Sangat Kurang Setuju	0	0	0	0	0,3925371
	Sangat Setuju	3	3	0	0	
	Setuju	0	0	0	0	
	Cukup Setuju	6	6	0	0	
	Kurang Setuju	4	0	4	0	1
	Sangat Kurang Setuju	5	0	5	0	

Pada Tabel 3 merupakan perhitungan pada iterasi ke-2. Dari Tabel 3, didapatkan bahwa faktor “Membaca Referensi” memiliki *gain* dengan nilai 1, “Mempraktikkan dari Lingkungan” memiliki *gain* dengan nilai 0,3925371, dan “Memanfaatkan Teknologi” memiliki *gain* dengan nilai 1. Dapat dilihat bahwa telah didapatkan *gain* yang sama yaitu pada faktor “Membaca Referensi” dan “Memanfaatkan Teknologi” dengan nilai *gain* adalah 1.

Jika *entropy* telah bernilai 0 pada salah satu Semesta, hal itu menandakan bahwa sudah memiliki *leaf* atau daun. Pada penelitian ini, penulis melakukan validasi pada *software* Rapidminer. Penulis menggunakan *decision tree* dan menghasilkan pohon keputusan sebagai berikut:



Gambar 1. Pohon keputusan pada *Rapidminer*.

Berdasarkan pohon keputusan di atas dihasilkan pohon keputusan peningkatan kemampuan Bahasa Inggris mahasiswa dalam bentuk *text view* yaitu, sebagai berikut:

Mendengar dari Lingkungan = 1.0: Tidak Mampu {Mampu=0, Tidak Mampu=9}

Mendengar dari Lingkungan = 2.0

| Membaca Referensi = 2.0: Tidak Mampu {Mampu=0, Tidak Mampu=3}

| Membaca Referensi = 4.0: Mampu {Mampu=2, Tidak Mampu=0}

| Membaca Referensi = 5.0: Tidak Mampu {Mampu=0, Tidak Mampu=2}

Mendengar dari Lingkungan = 3.0: Mampu {Mampu=9, Tidak Mampu=0}

Mendengar dari Lingkungan = 4.0

| Mempraktikkan dalam Lingkungan = 1.0: Tidak Mampu {Mampu=0, Tidak Mampu=4}

| Mempraktikkan dalam Lingkungan = 3.0: Tidak Mampu {Mampu=0, Tidak Mampu=7}

| Mempraktikkan dalam Lingkungan = 4.0: Mampu {Mampu=4, Tidak Mampu=0}

| Mempraktikkan dalam Lingkungan = 5.0: Mampu {Mampu=12, Tidak Mampu=0}

Mendengar dari Lingkungan = 5.0

| Memanfaatkan Teknologi = 1.0: Tidak Mampu {Mampu=0, Tidak Mampu=5}

| Memanfaatkan Teknologi = 2.0: Tidak Mampu {Mampu=0, Tidak Mampu=4}

| Memanfaatkan Teknologi = 3.0: Mampu {Mampu=6, Tidak Mampu=0}

| Memanfaatkan Teknologi = 5.0: Mampu {Mampu=3, Tidak Mampu=0}

Arti dari *Text View* di atas, yaitu:

- Faktor “Mendengar dari Lingkungan” yang “Sangat Kurang Setuju” tergolong “Tidak Mampu” karena terdapat 9 responden yang “Tidak Mampu” dan 0 responden yang “Mampu” terhadap Kemampuan Bahasa Inggris.
- Faktor “Mendengar dari Lingkungan” tergolong “Kurang Setuju”, jika:

- “Membaca Referensi” yang “Tidak Setuju” tergolong “Tidak Mampu” karena terdapat 3 responden yang “Tidak Mampu” dan 0 responden yang “Mampu”.
- “Membaca Referensi” yang “Setuju” tergolong “Mampu” karena terdapat 2 responden yang “Mampu” dan 0 responden yang “Tidak Mampu”.
- “Membaca Referensi” yang “Sangat Setuju” tergolong “Tidak Mampu” karena terdapat 2 responden yang “Tidak Mampu” dan 0 responden yang “Mampu”.
- Faktor “Mendengar dari Lingkungan” yang “Cukup Setuju” tergolong “Mampu” karena terdapat 9 responden yang “Mampu” dan 0 responden yang “Tidak Mampu” terhadap Kemampuan Bahasa Inggris.
- Faktor “Mendengar dari Lingkungan” tergolong “Setuju”, jika :
  - “Mempraktikkan dalam Lingkungan” yang “Sangat Kurang Setuju” tergolong “Tidak Mampu” karena terdapat 0 responden yang “Mampu” dan 4 responden yang “Tidak Mampu”.
  - “Mempraktikkan dalam Lingkungan” yang “Cukup Setuju” tergolong “Tidak Mampu” karena terdapat 0 responden yang “Mampu” dan 7 responden yang “Tidak Mampu”.
  - “Mempraktikkan dalam Lingkungan” yang “Setuju” tergolong “Mampu” karena terdapat 4 responden yang “Mampu” dan 0 responden yang “Tidak Mampu”.
  - “Mempraktikkan dalam Lingkungan” yang “Sangat Setuju” tergolong “Mampu” karena terdapat 12 responden yang “Mampu” dan 0 responden yang “Tidak Mampu”.
- Faktor “Mendengar dari Lingkungan” tergolong “Sangat Setuju”, jika :
  - “Memanfaatkan Teknologi” yang “Sangat Kurang Setuju” tergolong “Tidak Mampu” karena terdapat 0 responden yang “Mampu” dan 5 responden yang “Tidak Mampu”.
  - “Memanfaatkan Teknologi” yang “Kurang Setuju” tergolong “Tidak Mampu” karena terdapat 0 responden yang “Mampu” dan 4 responden yang “Tidak Mampu”.
  - “Memanfaatkan Teknologi” yang “Cukup Setuju” tergolong “Mampu” karena terdapat 6 responden yang “Mampu” dan 0 responden yang “Tidak Mampu”.
  - “Memanfaatkan Teknologi” yang “Sangat Setuju” tergolong “Mampu” karena terdapat 3 responden yang “Mampu” dan 0 responden yang “Tidak Mampu”.

Cara membaca pohon keputusan (*decision tree*) pada Gambar 1, yaitu: “Jika Mendengar dari Lingkungan = 1.0 maka Tidak Mampu. Jika Mendengar dari Lingkungan = 2.0 dan Membaca Referensi = 2.0 maka Tidak Mampu atau Membaca Referensi = 4.0 maka Mampu atau Membaca Referensi = 5.0 maka Tidak Mampu. Jika Mendengar dari Lingkungan = 3.0 maka Mampu. Jika Mendengar dari Lingkungan = 4.0 dan Mempraktikkan dalam Lingkungan = 1.0 maka Tidak Mampu atau Mempraktikkan dalam Lingkungan = 3.0 maka Tidak Mampu atau Mempraktikkan dalam Lingkungan = 4.0 maka Mampu atau Mempraktikkan dalam Lingkungan = 5.0 maka Mampu. Jika Mendengar dari Lingkungan = 5.0 dan Memanfaatkan Teknologi = 1.0 maka Tidak Mampu atau Memanfaatkan Teknologi = 2.0 maka Tidak Mampu atau Memanfaatkan Teknologi = 3.0 maka Mampu atau Memanfaatkan Teknologi = 5.0 maka Mampu”.

#### 4. Kesimpulan

Dari hasil penelitian yang diperoleh, dapat disimpulkan bahwa telah didapatkan faktor yang meningkatkan kemampuan Bahasa Inggris mahasiswa. Beberapa faktor yang digunakan dalam melakukan penelitian ini antara lain: Membaca Referensi (K1), Mendengar dari Lingkungan (K2), Mempraktikkan dalam Lingkungan (K3), dan Memanfaatkan Teknologi (K4). Hasil perhitungan dan validasi menyebutkan bahwa faktor yang mempengaruhi dalam meningkatkan kemampuan mahasiswa terhadap Bahasa Inggris adalah faktor “Mendengar dari Lingkungan” dengan proses perhitungan iterasi berhenti pada iterasi ke-2 dan *gain* dengan nilai 0,279838.

Hasil dari perhitungan penelitian ini menggunakan pengaplikasian Data Mining yaitu menggunakan Algoritma C4.5 yang di mana, dalam melakukan perhitungan untuk menentukan faktor yang meningkatkan Bahasa Inggris pada mahasiswa harus memperhatikan nilai *gain* yang paling tinggi dari faktor yang digunakan dalam penelitian ini. Pada Penelitian ini dilakukannya proses validasi dengan menggunakan *software* Rapidminer sehingga menghasilkan pohon keputusan (*decision tree*).

#### Daftar Pustaka

- [1] K. F. Irnanda and A. P. Windarto, “Penerapan Klasifikasi C4.5 Dalam Meningkatkan Kecakapan Berbahasa Inggris dalam Masyarakat,” in *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 2020, pp. 304–308.
- [2] F. Megawati, “Kesulitan Mahasiswa dalam Mencapai Pembelajaran Bahasa Inggris Secara Efektif,”



- Pedagog. J. Pendidik.*, vol. 5, no. 2, pp. 147–156, 2016.
- [3] W. Katrina, H. J. Damanik, F. Parhusip, D. Hartama, A. P. Windarto, and A. Wanto, “C.45 Classification Rules Model for Determining Students Level of Understanding of the Subject,” in *The International Conference on Computer Science and Applied Mathematic*, 2018, pp. 1–6, doi: 10.1088/1742-6596/1255/1/012005.
  - [4] M. Widyastuti, A. G. F. Simanjuntak, D. Hartama, A. P. Windarto, and A. Wanto, “Classification Model C.45 on Determining the Quality of Customer Service in Bank BTN Pematangsiantar Branch,” in *The International Conference on Computer Science and Applied Mathematic*, 2018, pp. 1–5, doi: 10.1088/1742-6596/1255/1/012002.
  - [5] Sudirman, A. P. Windarto, and A. Wanto, “Data mining tools | rapidminer: K-means method on clustering of rice crops by province as efforts to stabilize food crops in Indonesia,” in *2nd Nommensen International Conference on Technology and Engineering*, 2018, pp. 1–8, doi: 10.1088/1757-899X/420/1/012089.
  - [6] R. W. Sari, A. Wanto, and A. P. Windarto, “Implementasi Rapidminer dengan Metode K-Means (Study Kasus: Imunisasi Campak pada Balita Berdasarkan Provinsi),” in *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 2018, vol. 2, pp. 224–230.
  - [7] N. Rofiqo, A. P. Windarto, and D. Hartama, “Penerapan Clustering pada Penduduk yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means,” in *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 2018, vol. 2, pp. 216–223.
  - [8] M. G. Sadewo, A. P. Windarto, and A. Wanto, “Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 311–319, 2018, doi: 10.30865/komik.v2i1.943.
  - [9] D. Hartama, A. P. Windarto, and A. Wanto, “The Application of Data Mining in Determining Patterns of Interest of High School Graduates,” in *The International Conference on Computer Science and Applied Mathematic*, 2019, pp. 1–6, doi: 10.1088/1742-6596/1339/1/012042.
  - [10] S. Haryati, A. Sudarsono, and E. Suryana, “Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu),” *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
  - [11] A. S. Febriarini and E. Z. Astuti, “Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang,” *J. Eksplora Inform.*, vol. 8, no. 2, pp. 95–103, 2019, doi: 10.30864/eksplora.v8i2.156.
  - [12] J. Han, M. Kamber, and J. Pei, *Data Mining : Concepts and Techniques Third Edition*, 3rd ed. Elsevier, 2012.
  - [13] Y. Yulia and A. D. Putri, “Data Mining Menggunakan Algoritma C4.5 untuk Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen di Kota Batam,” *Comput. Based Inf. Syst. J.*, vol. 7, no. 2, pp. 56–66, 2019.
  - [14] D. A. Silitonga, A. P. Windarto, and D. Hartama, “Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi,” in *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*, 2019, pp. 313–318.
  - [15] D. Armiadi, “Prediksi Kepuasan Stakeholder Dengan Algoritma C.45,” *Lentera J. Ilm. Sains dan Teknol.*, vol. 15, no. 16, pp. 69–76, 2015.